Deviation inequality for monotonic Boolean functions with application to a number of k-cycles in a random graph.

Dmitry Panchenko
Massachusetts Institute of Technology

Abstract

Using Talagrand's concentration inequality on the discrete cube $\{0,1\}^m$ we show that given a real-valued function $Z(x)$ on $\{0,1\}^m$ that satisfies certain monotonicity conditions one can control the deviations of $Z(x)$ above its median by a local Lipschitz norm of Z at the point x. As one application, we give a simple proof of a nearly optimal deviation inequality for the number of k-cycles in a random graph.

1 Introduction and main results.

In this paper we suggest a new way to use Talagrand's concentration inequality on the cube to control the deviations of Boolean functions that satisfy certain monotonicity conditions. As one application we prove a suboptimal deviation inequality for the count of k-cycles in a random graph.

Let $\mathcal{X} = \{0,1\}$ and define a probability measure μ on $\mathcal{X}$ by $\mu(\{1\}) = p, \mu(\{0\}) = 1-p$. Consider a product space $\mathcal{X}^m$ with a product probability measure $\mathbb{P} = \mu^m$. Given a function $Z : \mathcal{X}^m \to \mathbb{R}$ and a point $x = (x_1, \ldots, x_m) \in \mathcal{X}^m$ we define

$$V_i(x) = Z(x) - Z(x_1, \ldots, x_{i-1}, 0, x_{i+1}, \ldots, x_m),$$

and

$$V(x) = \sum_{i=1}^{m} V_i^2(x).$$

Note that $V_i(x) = 0$ if $x_i = 0$.

Let us state the main result of this paper.

Theorem 1 *If $Z(x)$ and $V_i(x), i \le m$ are non-decreasing in each coordinate then for any $a \in \mathbb{R}$ and $t > 0$,*

$$\mathbb{P}\big(Z(x) \ge a + \sqrt{V(x)t}\big)\mathbb{P}\big(Z(x) \le a\big) \le e^{-t/2}. \tag{1.1}$$

To understand the statement of Theorem 1, we notice that a function $(V(x))^{1/2}$ can be interpreted as a kind of discrete Lipschitz norm of Z locally at the point x. For example, since $Z(x)$ is defined only on the vertices of m-dimensional cube, if one extends $Z(x)$ linearly from the point x to its neighbours only along the coordinates where $x_i = 1$, then $(V(x))^{1/2}$ is the norm of that linear map. Indeed, if we denote the map by $L : \mathbb{R}^m \to \mathbb{R}$, then

$$L(z) = \sum_{i=1}^{m} V_i(x)(z_i - x_i) + Z(x),$$

and $\|L\| = (V(x))^{1/2}$.

The proof of Theorem1 is based on Talagrand's concentration inequality on $\mathcal{X}^m$. In order to give more clear interpretation of (1.1), let us compare it with some typical ways of using Talagrand's inequality. One common application is the following. Given a convex function $f : [0,1]^m \to \mathbb{R}$ with a Lipschitz norm

$$\|f\|_L = \sup_{x,y\in[0,1]^m} \frac{|f(x) - f(y)|}{|x-y|} < \infty,$$

where the supremum is taken over $x \neq y$, the following inequality holds:

$$\mathbb{P}\big(f \geq a + \|f\|_L \sqrt{t}\big)\mathbb{P}\big(f \leq a\big) \leq e^{-t/2}. \tag{1.2}$$

If the function f is defined only on the vertices of the cube $\mathcal{X}^m$, then it is possible to state a similar result where one has to use a discrete analog of the Lipschitz norm. For example, the following deviation inequality holds (see [1]). For $i \leq m$ we define $x^i \in \mathcal{X}^m$ such that $x^i_j = x_j$ for $j \neq i$ and $x^i_i \neq x_i$, and define

$$\|f\|_d = \sup_{x\in\mathcal{X}^m} \Big(\sum_{i=1}^{m} (f(x) - f(x^i))^2\Big)^{1/2}.$$

Then

$$\mathbb{P}\big(f \geq \mathbb{E}f + \|f\|_d \sqrt{t}\big) \leq e^{-t/4}. \tag{1.3}$$

Both inequalities (1.2) and (1.3) use global Lipschitz condition to control the deviation of $f(x)$. Theorem 1 suggests a possibility of using a local Lipschitz norm $V(x)^{1/2}$ at the point x, provided that the monotonicity conditions are satisfied. The reason why we compute the Lipschitz norm only in the direction of decreasing Z is because we control the deviation of Z *above* level a. Theorem 1 is similar in spirit to the ideas in [6], [8],[9] (see also references therein), where the authors describe a way of using average Lipschitz norm of Z to control its deviations.

One example when the monotonicity conditions are satisfied is the following. Let us consider a set of indices $\mathcal{M} = \{1, 2, \ldots, m\}$ and a set of nonnegative numbers $\alpha_{\mathcal{C}} \geq 0$ indexed by the subsets $\mathcal{C} \subseteq \mathcal{M}$. Consider the function $Z(x)$ defined by

$$Z(x) = \sum_{\mathcal{C}\subseteq\mathcal{M}} \alpha_{\mathcal{C}} \prod_{i\in\mathcal{C}} x_i. \tag{1.4}$$

In this case the fact that $\alpha_{\mathcal{C}}$'s are non-negative implies that functions $Z(x)$ and $V_i(x), i \leq m$ are non-decreasing in each coordinate. Below we will consider the example of counting the number of k-cycles in a random graph which can be represented in the form (1.4) and, thus, Theorem 1 is applicable.

Consider a standard Erdös-Rényi model of a random graph $G(n,p)$. Let V be a set of n vertices, $m = \binom{n}{2}$ and let $E = \{e_1, \ldots, e_m\}$ denote a set of edges of a complete graph K_n on n vertices. Given $x = (x_1, \ldots, x_m) \in \mathcal{X}^m$, the fact that $x_i = 1$ or 0 describes that the edge e_i is present or not present in the graph $G(n,p)$ respectively. Let

$$C_k = \Big\{ \{e_{i_1}, \ldots, e_{i_k}\} : e_{i_j} \in E, \text{ and } \{e_{i_1}, \ldots, e_{i_k}\} \text{ form a } k\text{-cycle} \Big\}$$

be a collection of all k-cycles, and for $e \in E$ let

$$C_k(e) = \{c \in C_k : e \in c\}$$

be a set of all k-cycles containing the edge e. We consider the following function on $\mathcal{X}^m$

$$Z(x) = \sum_{c \in C_k} \prod_{e \in c} x_e,$$

which is the number of k-cycles in a random graph $G(n,p)$. In this case $V(x)$ can be clearly written as

$$V(x) = \sum_{e \in E} x_e \Big(\sum_{c \in C_k(e)} \prod_{e' \neq e} x_{e'} \Big)^2. \tag{1.5}$$

In this case, in order to use Theorem 1 to control the deviation of $Z(x)$ above its median $M(Z)$ (or its expectation $\mathbb{E}Z$) we will proceed by showing how to control $V(x)$ in terms of $Z(x)$.

We assume that for some large enough $C(k) > 0$,

$$np \geq C(k) \log n. \tag{1.6}$$

The following theorem holds.

Theorem 2 *If (1.6) holds then there exists a constant $C(k) > 0$ that depends on k only such that*

$$\mathbb{P}\Big(V(x) \geq C(k)\big((np)^{k-2} Z(x) + (np)^{2(k-1)} \big) \Big) \leq \exp\Big(-\frac{(np)^2}{C(k) \log\log np} \Big).$$

Theorems 1 and 2 will readily imply the following theorem.

Theorem 3 *If (1.6) holds then,*

(1) For any $\varepsilon > 0$ there exists a constant $C(k, \varepsilon)$ that depends on k and ε only such that the following holds

$$\text{If} \quad \mathbb{E}Z \geq C(k,\varepsilon) \quad \text{then} \quad M(Z) \leq (1+\varepsilon)\mathbb{E}Z.$$

(2) There exists a constant $C(k) > 0$ such that if $\mathbb{E}Z \geq C(k)$ then

$$\mathbb{P}(Z \geq 2\mathbb{E}Z) \leq \exp\Big(-\frac{(np)^2}{C(k) \log\log np} \Big). \tag{1.7}$$

Recently the authors of [5] proved a more general result describing the deviations of the count of any subgraph in a random graph. In the case of k-cycles their bound gives

$$\mathbb{P}(Z \geq 2\mathbb{E}Z) \leq \exp(-(np)^2/C(k)).$$

This shows that the factor $\log\log np$ in (1.7) is unnecessary, but at the moment we don't see how to get rid of it using our approach. This has nothing to do with Theorem 1, since the factor $\log\log np$ comes directly from Theorem 2 which, probably, can be improved. In the case of triangles ($k=3$) the bound $\mathbb{P}(Z \geq 2\mathbb{E}Z) \leq \exp(-(np)^2/C(k))$ was also proved in [6].

2 Proof of Theorem 1.

Talagrand's concentration inequality on the discrete cube is the main tool in the proof of Theorem 1. Let us recall it first.

Given a point $x \in \mathcal{X}^m = \{0,1\}^m$ and a set $\mathcal{A} \subseteq \mathcal{X}^m$, let us denote

$$U_{\mathcal{A}}(x) = \{(s_i)_{i \leq m} \in \{0,1\}^m, \exists y \in \mathcal{A}, s_i = 0 \Rightarrow y_i = x_i\}.$$

The "convex hull" distance between the point x and the set $\mathcal{A}$ is defined as

$$f_c(\mathcal{A}, x) = \inf\{|s| : s \in \mathrm{conv}U_{\mathcal{A}}(x)\},$$

where $|s|$ is the Euclidean norm of s. The concentration inequality of Talagrand (Theorem 4.3.1 in [7]) states the following.

Proposition 1 *For any $t > 0$,*

$$\mathbb{P}(\mathcal{A})\mathbb{P}\Big(x \in \mathcal{X}^m : f_c^2(\mathcal{A}, x) \geq t\Big) \leq e^{-t/2}. \tag{2.1}$$

The main feature of this distance is that (Theorem 4.1.2 in [7])

$$\forall (\lambda_i)_{i \leq m} \quad \exists y \in \mathcal{A} \qquad \sum_{i=1}^m \lambda_i I(y_i \neq x_i) \leq f_c(\mathcal{A}, x)\Big(\sum_{i=1}^m \lambda_i^2\Big)^{1/2}. \tag{2.2}$$

Proof of Theorem 1. For a fixed number $a \in \mathbb{R}$ consider a set

$$\mathcal{A} = \{y \in \mathcal{X}^m : Z(y) \leq a\}.$$

For a fixed $x \in \mathcal{X}^m$ and an arbitrary $y \in \mathcal{A}$, since $Z(y) \leq a$, we can write $Z(x) - a \leq Z(x) - Z(y)$. Consider three sets of indices

$$I_1 = \{i : x_i = 1, y_i = 0\}, \quad I_2 = \{i : x_i = 0, y_i = 1\}, \quad I_3 = \{i : x_i = y_i\}.$$

Without loss of generality we will assume that $I_1 = \{1, \ldots, k\}$ and $I_2 = \{k+1, \ldots, l\}$. Define a sequence

$$z^i = (y_1, \ldots, y_i, x_{i+1}, \ldots, x_m), \ \ i = 0, \ldots, m.$$

We have

$$Z(x) - Z(y) = \sum_{i=1}^{m} (Z(z^{i-1}) - Z(z^i)) = \sum_{i=1}^{m} (Z(z^{i-1}) - Z(z^i)) I(x_i \neq y_i),$$

since $x_i = y_i$ (i.e. $i > l$) implies that $Z(z^{i-1}) - Z(z^i) = 0$. We have

$$Z(z^{i-1}) - Z(z^i) = 0 \leq V_i(x) \quad \text{for} \quad i = l+1, \ldots, m,$$

since for this range of indices $z^{i-1} = z^i$,

$$Z(z^{i-1}) - Z(z^i) \leq 0 \leq V_i(x) \quad \text{for} \quad i = k+1, \ldots, l,$$

since the function Z is non-decreasing in each coordinate and for $i \in I_2$, $z_i^{i-1} = 0$, $z_i^i = 1$ and all other coordinates of z^{i-1} and z^i coincide, and

$$Z(z^{i-1}) - Z(z^i) = V_i(z^{i-1}) \leq V_i(x) \quad \text{for} \quad i = 1, \ldots, k,$$

since for $i \in I_1$ each coordinate of z^{i-1} is smaller than the corresponding coordinate of x. Thus we proved that for any $y \in \mathcal{A}$

$$Z(x) - a \leq \sum_{i=1}^{m} V_i(x) I(x_i \neq y_i).$$

By (2.2) there exists $y \in \mathcal{A}$ such that the last expression can be bounded

$$\sum_{i=1}^{m} V_i(x) I(x_i \neq y_i) \leq f_c(\mathcal{A}, x) \Big(\sum_{i=1}^{m} V_i^2(x) \Big)^{1/2} = f_c(\mathcal{A}, x) \sqrt{V(x)}$$

Talagrand's inequality (2.1) states that

$$\mathbb{P}(f_c(\mathcal{A}, x) \geq \sqrt{t}) \mathbb{P}(\mathcal{A}) \leq e^{-t/2},$$

and, therefore, we finally get

$$\mathbb{P}(Z(x) \geq a + \sqrt{V(x) t}) \mathbb{P}(Z(x) \leq a) \leq e^{-t/2}.$$

□

3 Proof of Theorem 2.

We will denote by d_v the degree of a vertex v. Consider the sequence of sets

$$V_1 = \{v : d_v < 16np\}, \quad V_j = \{v : d_v \in [2^{j+2}np, 2^{j+3}np)\}, \quad j \geq 2. \tag{3.1}$$

We will start by stating several basic facts that will be used in the proof of Theorem 2.

Lemma 1 *If (1.6) holds then,*

$$\mathbb{P}\Big(\exists v : d_v \geq (np)^2\Big) \leq e^{-(np)^2/2}. \tag{3.2}$$

Proof. For a fixed vertex v its degree d_v is a sum of $(n-1)$ independent variables with the distribution μ. Using Bernstein's inequality one can easily check that for $np \geq 4$

$$\mathbb{P}\Big(d_v \geq (np)^2\Big) \leq e^{-(np)^2}.$$

The union bound will produce a factor n and, therefore, using (1.6) implies (3.2).

□

Thus, with high probability we can assume that the degree of each vertex is bounded by $(np)^2$ and, therefore, we can only consider the sets V_j in (3.1) such that $2^{j+2} \leq np$ and, therefore, $j \leq \log np$.

Next we will bound the cardinality of each V_j.

Lemma 2 *For $C(k) > 0$ large enough we have,*

$$\mathbb{P}\Big(\exists 2 \leq j \leq \log np \quad \mathrm{card} V_j \geq \frac{np}{j 2^j \log\log np}\Big) \leq \exp\Big(-\frac{(np)^2}{C(k)\log\log np}\Big). \tag{3.3}$$

Proof. We copy the proof from [6] (see equation (12) in section 4.2 there). For a fixed $j \geq 2$, assume that

$$\mathrm{card} V_j \geq r = \frac{np}{j 2^j \log\log np}.$$

In this case, there exists a set of r vertices each with degree at least $2^{j+2}np$. It implies that the number of edges containing exactly one of these vertices exceeds

$$(2^{j+2}np - r)r \geq 2^{j+1}npr,$$

The probability that such a set of edges exists is bounded by

$$\binom{n}{r}\binom{r(n-r)}{2^{j+1}npr} p^{2^{j+1}npr} \leq \exp\Big(r\log\frac{en}{r} + 2^{j+1}npr\log\frac{er(n-r)}{2^{j+1}npr} + 2^{j+1}npr\log p\Big)$$
$$\leq \exp\Big(r\log\frac{en}{r} + 2^{j+1}npr\log\frac{e}{2^{j+1}}\Big) \leq \exp\Big(-\frac{(np)^2}{C(k)\log\log np}\Big),$$

where in the last inequality we used the estimate

$$2^{j+1}npr\log\frac{e}{2^{j+1}} \leq \frac{(np)^2}{2j\log\log np}\log\frac{1}{2^{j-1}} \leq -\frac{(np)^2}{C(k)\log\log np},$$

and the first term $r\log(en/r)$ was negligible compared to the second term. Taking the union bound over $j \leq \log np$, we get a factor $\log np$ in front of the exponent that can be ignored by increasing $C(k)$.

□

Before we will state our next lemma, we need to make one remark about the proof of Theorem 2. Multiplying out the right-hand side of (1.5) we observe that $V(x)$ can be written as a sum of terms

$$x_e \prod_{e' \in c} x_{e'} \prod_{e' \in c'} x_{e'}, \text{ where } e \in E, c, c' \in C_k(e).$$

Each of these terms may appear several times, but, clearly, the number of appearances will be bounded by $C(k)$ that depends on k only. Each of these term represents two cycles that have at least one edge in common. There are many different isometric configurations of such two cycles but, clearly, the number of them is bounded by a constant that depends on k only. Hence, $V(x)$ can be decomposed into the sum of the counts of such pairs of cycles over different configuration. With minor modifications it is possible to prove the statement of the theorem for each of these configuration.

We will only look at the pairs of cycles that have exactly one edge in common. Let us denote the number of such pair by $W(x)$. We will identify each pair of cycles with an injection $\sigma : \{1, \ldots, 2k-2\} \to V(G)$, such that $(\sigma(1), \sigma(2) \ldots, \sigma(k))$ and $(\sigma(k), \sigma(k+1), \ldots, \sigma(1))$ are the ordered vertices of these two cycles, and $\sigma(1)\sigma(k)$ is their only common edge. Let us denote the set of these injections by Σ_0.

Lemma 3 *There is a partition of vertices* $V(G) = F_1 \cup \ldots \cup F_{2k-2}$ *such that*

$$W(x) \le C(k) \text{card}\{\sigma \in \Sigma_0 : \forall i \ \ \sigma(i) \in F_i\}. \tag{3.4}$$

Proof. See Proposition 1.3 in [3].

□

Let us denote the set in the statement of Lemma 3 by

$$\Sigma = \{\sigma \in \Sigma_0 : \forall i \ \ \sigma(i) \in F_i\}. \tag{3.5}$$

Proof of Theorem 2. By Lemma 3 and the discussion preceeding Lemma 3, all we need to do is to estimate the cardinality of Σ in (3.5). Let us consider the event

$$\mathcal{E} = \Big\{\forall j \ge 2 \ \ \text{card} V_j \le \frac{np}{j 2^j \log\log np}\Big\} \bigcup \big\{\forall v : d_v \le (np)^2\big\}.$$

By Lemma 1 and Lemma 2 this event holds with probability at least

$$1 - \exp\Big(-\frac{(np)^2}{C(k)\log\log np}\Big),$$

for some $C(k)$ large enough. From now on we assume that this event occurs. For each vertex $v \in F_1$ let us denote

$$S_l(v) = \{(\sigma(1), \ldots, \sigma(l)) : \sigma \in \Sigma, \sigma(1) = v\}, \ \ l \ge 1.$$

Let us denote $d_v^+ = \max(d_v, np)$. We will prove that if $\mathcal{E}$ occurs than for $l \ge 2$

$$\text{card} S_l(v) \le C(k) d_v^+ (np)^{l-2}. \tag{3.6}$$

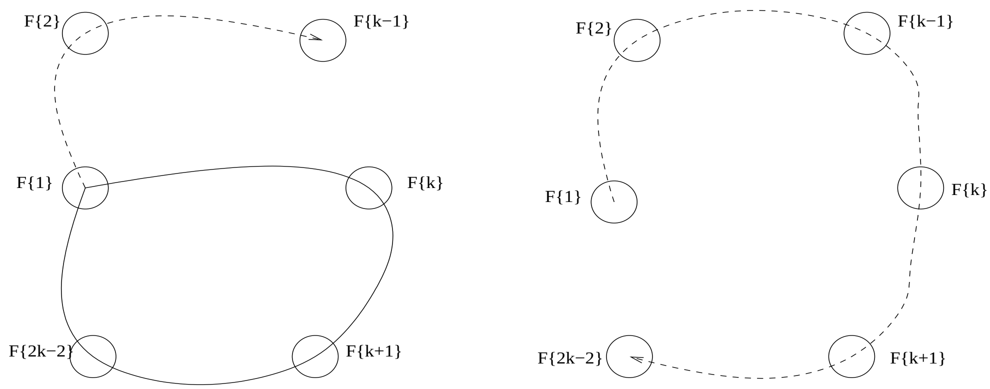

Figure 1: (1) Counting $\sigma \in \Sigma_1$. (2) Counting $\sigma \in \Sigma_2$.

For $l = 2$, this obviously holds with $C(k) = 1$. We proceed by induction over l. Let us decompose

$$S_l(v) = \bigcup_{j \geq 1} S_l^j(v),$$

where

$$S_l^j(v) = \{(\sigma(1), \ldots, \sigma(l)) \in S_l(v) : \sigma(l-1) \in V_j\}.$$

To bound the cardinality of $\mathrm{card} S_l^1(v)$ we use the induction hypothesis and the fact that the degree of each vertex in the set V_1 is bounded by $16np$. We get

$$\mathrm{card} S_l^1(v) \leq (16np)\ \mathrm{card} S_{l-1}(v) \leq C(k) d_v^+ (np)^{l-2}.$$

To bound the cardinality of $S_l^j(v)$ for $j \geq 2$ we notice that for each $(\sigma(1), \ldots, \sigma(l)) \in S_l^j(v)$, we have $(\sigma(1), \ldots, \sigma(l-2)) \in S_{l-2}(v)$ and $\sigma(l-1) \in V_j$. On the event $\mathcal{E}$ we can control the cardinality of V_j and, moreover, the degree $d_{\sigma(l-1)} \leq 2^{j+3} np$. For $l = 3$, since $\mathrm{card} S_1(v) = 1$, we get

$$\mathrm{card} S_3^j(v) \leq (\mathrm{card} V_j)\ (2^{j+3} np) \leq \frac{np}{2^j j \log\log np} 2^{j+3} np \leq 8(np)^2 \frac{1}{j \log\log np}$$

and since on the event $\mathcal{E}$ we can assume that $2^{j+2} \leq np$, which implies that $j \leq \log np$, we get

$$\sum_{j=2}^{\log np} \mathrm{card} S_3^j(v) \leq 8(np)^2 \sum_{j=2}^{\log np} \frac{1}{j \log\log np} \leq C(k)(np)^2 \leq C(k) d_v^+ (np),$$

and this proves the induction step for $l = 3$. The last inequality explains the appearance of the factor $\log\log np$ in Theorem 3. Similarly, for $l \geq 4$ we get

$$\begin{aligned} &\mathrm{card} S_l^j(v) \leq \mathrm{card} S_{l-2}(v)\ (\mathrm{card} V_j)\ (2^{j+3} np) \\ &\leq C(k) d_v^+ (np)^{l-4} \frac{np}{2^j j \log\log np} 2^{j+3} np \leq C(k) d_v^+ (np)^{l-2} \frac{1}{j \log\log np} \end{aligned}$$

and

$$\sum_{j=2}^{\log np} \mathrm{card} S_l^j(v) \le C(k) d_v^+(np)^{l-2} \sum_{j=2}^{\log np} \frac{1}{j \log\log np} \le C(k) d_v^+(np)^{l-2}.$$

This completes the proof of the induction step and (3.6).

To estimate the cardinality of Σ we will decompose it into $\Sigma = \Sigma_1 \cup \Sigma_2$, where

$$\Sigma_1 = \{\sigma \in \Sigma : \sigma(1) \in V_1\}, \quad \Sigma_2 = \{\sigma \in \Sigma : \sigma(1) \in V_j, 2 \le j\}.$$

We will estimate the cardinality of Σ_1 and Σ_2 differently (the idea is illustrated in Figure 1). First of all, since we can control the cardinality of V_j for $j \ge 2$, we will simply use (3.6) for $l = 2k-2$ to compute the number of different paths from $v \in F_1$ to F_{2k-2}, and then add them up. This will give us the bound on cardinality of Σ_2. On the other hand, for $\sigma \in \Sigma_1$ we can represent it as a cycle on $F_1, F_k, \ldots, F_{2k-2}$ and a path from F_1 to F_{k-1}. In this case, the number of cycles is bounded by $Z(x)$, and to bound the number of paths we again use (3.6).

Let first estimate the cardinality of Σ_2. First of all by (3.6) for each vertex $v \in V_j \cap F_1$

$$\mathrm{card} S_{2k-2}(v) \le C(k) 2^{j+3}(np)(np)^{2k-4},$$

and, therefore, on the event $\mathcal{E}$,

$$\mathrm{card}\{\sigma \in \Sigma : \sigma(1) \in V_j\} \le C(k)2^{j+3}(np)^{2k-3}(\mathrm{card} V_j) \le C(k)2^{j+3}(np)^{2k-3}\frac{np}{j2^j \log\log np}.$$

When we add up these injections over $j \ge 2$ we get

$$\sum_{j=2}^{\log np} C(k)2^{j+3}(np)^{2k-3}\frac{np}{j2^j \log\log np} \le C(k)(np)^{2k-2}.$$

This accounts for the second term in the bound of the theorem.

Now consider all injections σ such that $\sigma(1) \in V_1$. Consider the trace of the set of images of the injections from Σ (in other words, pairs of cycles) on the set

$$(V_1 \cap F_1) \cup F_k \cup F_{k+1} \cup \ldots \cup F_{2k-2},$$

i.e.

$$\mathcal{P} = \{(\sigma(1), \sigma(k), \sigma(k+1), \ldots, \sigma(2k-2)) : \sigma \in \Sigma, \sigma(1) \in V_1\}.$$

First of all, the cardinality of $\mathcal{P}$ is bounded by $Z(x)$, since $\mathcal{P}$ can be identified with the subset of all cycles in the random graph. Moreover, for each $(v_1, v_k, v_{k+1}, \ldots, v_{2k-2}) \in \mathcal{P}$, the number of injections $\sigma \in \Sigma$ such that $\sigma(1) = v_1, \sigma(k) = v_k, \ldots, \sigma(2k-2) = v_{2k-2}$ is bounded by $\mathrm{card} S_{k-1}(v_1)$, since all values of the injection are fixed except for $\sigma(2), \ldots, \sigma(k-1)$. But since $v_1 \in V_1$ implies that the degree $d_{v_1} \le 16np$ and, thus, $d_{v_1}^+ \le 16np$, we have by (3.6)

$$\mathrm{card} S_{k-1}(v_1) \le C(k)(np)(np)^{k-3} \le C(k)(np)^{k-2}.$$

Therefore, the cardinality of all injections such that $\sigma(1) \in V_1$ is bounded by

$$C(k)(\mathrm{card}\mathcal{P})(np)^{k-2} \le C(k)Z(x)(np)^{k-2},$$

which accounts for the first term in the statement of the theorem.

□

4 Proof of Theorem 3.

Theorem 2 implies that for any $a \in \mathbb{R}$ and $t > 0$,

$$\mathbb{P}\Big(Z \le a + \sqrt{Vt}\Big) \le \mathbb{P}\Big(Z \le a + \Big(C(k)\big((np)^{k-2}Z + (np)^{2(k-1)}\big)t\Big)^{1/2}\Big)$$
$$+\mathbb{P}\Big(V \ge C(k)((np)^k Z + (np)^{2k})\Big)$$
$$\le \mathbb{P}\Big(Z \le a + \Big(C(k)\big((np)^{k-2}Z + (np)^{2(k-1)}\big)t\Big)^{1/2}\Big) + \exp\Big(-\frac{(np)^2}{C(k)\log\log np}\Big),$$

which implies that

$$\mathbb{P}\Big(Z \ge a + \Big(C(k)\big((np)^{k-2}Z + (np)^{2(k-1)}\big)t\Big)^{1/2}\Big) \;\le\; \mathbb{P}\Big(Z \ge a + \sqrt{Vt}\Big) \;+\; \exp\Big(-\frac{(np)^2}{C(k)\log\log np}\Big).$$

Multiplying both sides by $\mathbb{P}(Z \le a)$ and using Theorem 1 with $t = 2\varepsilon(np)^2$ we get

$$\mathbb{P}\Big(Z \ge a + \Big(C(k)\varepsilon\big((np)^k Z + (np)^{2k}\big)\Big)^{1/2}\Big)\mathbb{P}\Big(Z \le a\Big)$$
$$\le \mathbb{P}(Z \le a)\exp\Big(-\frac{(np)^2}{C(k)\log\log np}\Big) + e^{-\varepsilon(np)^2}. \tag{4.1}$$

If we take

$$a = M - \sqrt{C(k)\varepsilon((np)^k M + (np)^{2k})} \tag{4.2}$$

then, clearly, the following two events are equal

$$\{Z \ge a + \sqrt{C(k)\varepsilon((np)^k Z + (np)^{2k})}\} = \{Z \ge M\},$$

and, therefore, (4.1) implies that for $np \ge C(k)$ for large enough $C(k) > 0$,

$$\mathbb{P}\Big(Z \le M - \sqrt{C(k)\varepsilon((np)^k M + (np)^{2k})}\Big) \le 3e^{-\varepsilon(np)^2}.$$

Since $\mathbb{E}Z \ge a\mathbb{P}(Z \ge a)$, the choice of a as in (4.2) gives

$$M - \sqrt{C(k)\varepsilon((np)^k M + (np)^{2k})} \le \mathbb{E}Z\Big(1 - 3e^{-\varepsilon(np)^2}\Big)^{-1}.$$

This, clearly, implies the first statement of Theorem 3.

To prove the second statement we use (4.1) with $a = M$, and assume that np is large enough, so that $M \le (1+\varepsilon)\mathbb{E}Z$. Then with probability at least

$$1 - 2e^{-\varepsilon(np)^2} - \exp\Big(-\frac{(np)^2}{C(k)\log\log np}\Big)$$

we have

$$Z \le (1+\varepsilon)\mathbb{E}Z + \sqrt{C(k)\varepsilon((np)^k Z + (np)^{2k})}.$$

Since $\mathbb{E}Z \sim (np)^k$, for small enough ε this implies that $Z \le 2\mathbb{E}Z$, which completes the proof of the second statement of Theorem 3.

$\square$

Acknowledgment. We want to thank anonymous referee for helpful comments, especially, for suggesting the present formulation of Theorem 1.

References

[1] S. Bobkov, On the Gross and Talagrand inequalities on the discrete cube, *Vestn. Syktyvkar Univ. Ser.1 Mat. Mekh. Inform.* (1995), no.1, 12 -19.

[2] S. Boucheron, G. Lugosi, P. Massart, Concentration inequalities using the entropy method, To appear in *Ann. Probab.* (2002).

[3] E. Friedgut, J. Kahn, On the number of copies of one hypergraph in another, *Israel Journal of Mathematics* **105** (1998), 251 - 256.

[4] S. Janson, A. Ruciński, The infamous upper tail. Probabilistic methods in combinatorics, *Random Structures Algorithms* **20** (2002), no.3, 317 - 342.

[5] S. Janson, K. Oleszkiewicz, A. Ruciński, Upper tail for subgraph counts in random graphs, preprint (2002).

[6] J.H. Kim, V.H. Vu, Divide and conquer martingales and the number of triangles in a random graph, preprint (2002).

[7] M. Talagrand, Concentration of measure and isoperimetric inequalities in product spaces, *Publications Mathématiques de l'I.H.E.S.* **81** (1995), 73-205.

[8] V.H. Vu, A large deviation result on the number of small subgraphs of a random graph, *Combin. Probab. Comput.* **10** (2001), no.1, 79 - 94.

[9] V.H. Vu, Concentration of non-Lipschitz functions and applications. Probabilistics methods in combinatorial optimization, *Random Structures Algorithms,* **20** (2002), no.3, 262 - 316.

Department of Mathematics
Massachusetts Institute of Technology
77 Massachusetts Avenue, Room 2-181
Cambridge, MA, 02139-4307
URL: http://www-math.mit.edu/˜panchenk e-mail: panchenk@math.mit.edu